Prediction of two periods by simple autoregressive models with one lag

den Butter, F.A.G.

published in
Economist
1975

DOI (link to publisher)
10.1007/BF02115563

PREDICTION OF TWO PERIODS BY SIMPLE AUTOREGRESSIVE MODELS WITH ONE LAG

BY

F. A. G. DEN BUTTER*

1 THE PROBLEM

For purposes of prediction endogenous variables must be expressed in terms of past values and predetermined variables, as e.g. in the reduced form of large econometric models. Simple equations of this type are

$$\boldsymbol{Y}_t = \alpha + \beta \boldsymbol{Y}_{t-1} + \boldsymbol{U}_t \tag{1}$$

and

$$\boldsymbol{Y}_t = \alpha + \beta \boldsymbol{Y}_{t-1} + \gamma X_t + \boldsymbol{U}_t \tag{2}$$

where Y is endogenous, X predetermined (and perfectly predictable) and $\boldsymbol{U}_t$ a random variable with[1]

$$E(\boldsymbol{U}_t) = 0, \qquad E(\boldsymbol{U}_t^2) = \sigma^2; \tag{3}$$

$$E(\boldsymbol{U}_t \boldsymbol{U}_{t'}) = 0 \quad \text{for all } t \neq t'.$$

Prediction usually proceeds by analogy, so that in the case of the autoregressive model (1) (AR(1) model in Box and Jenkins [1] notation) the *predictors* of Y one and two periods ahead of a starting value $\boldsymbol{Y}_\tau$ are

$$\hat{\boldsymbol{Y}}_{\tau+1} = \boldsymbol{a} + \boldsymbol{b}\boldsymbol{Y}_\tau \tag{4}$$

$$\hat{\boldsymbol{Y}}_{\tau+2} = \boldsymbol{a} + \boldsymbol{b}\hat{\boldsymbol{Y}}_{\tau+1}$$

$$= \boldsymbol{a} + \boldsymbol{a}\cdot\boldsymbol{b} + \boldsymbol{b}^2\boldsymbol{Y}_\tau. \tag{5}$$

* I am indebted to Dr. M. M. G. Fase and professor A. H. Q. M. Merkies for useful suggestions and remarks and above all to professor J. S. Cramer for his critical comments and his great help in drafting this article. Currently I am research associate of the Econometric and Special Studies Section of the Domestic Research Department at De Nederlandsche Bank N.V., P.O. Box 98, Amsterdam, but the main part of this study was completed at the Instituut voor Actuariaat en Econometrie of the University of Amsterdam.

1 Random variables are displayed in bold type.

Here $\boldsymbol{a}$ and $\boldsymbol{b}$ are any estimators of α and β, derived from the observed time series Y_t, $t = 0 \ldots T$, and random because these refer to a variable that is random according to (1).

The properties of various functions $\boldsymbol{a}$ and $\boldsymbol{b}$ as estimators of α and β have often been investigated but less attention has so far been paid to their performance as predictors in (4) and (5). The two may easily conflict. In order to show this we define the *predictor error*

$$\boldsymbol{f}_t = \hat{\boldsymbol{Y}}_t - \boldsymbol{Y}_t, \tag{6}$$

and consider its expectation for unbiased estimators

$$E(\boldsymbol{a} - \alpha) = 0, \quad E(\boldsymbol{b} - \beta) = 0. \tag{7}$$

For prediction one period ahead we find

$$E(\boldsymbol{f}_{\tau+1}) = \operatorname{cov}(\boldsymbol{b}, \boldsymbol{Y}_\tau), \tag{8}$$

provided only $E(\boldsymbol{Y}_\tau)$ is finite. This will hardly ever be zero, for $\boldsymbol{Y}_\tau$ has been generated by the same process (1) as the observations $\boldsymbol{Y}_t$, $t = 0 \ldots T$, so that $\boldsymbol{Y}_\tau$ and $\boldsymbol{b}$ cannot be strictly independent. But (8) is zero and the predictor unbiased if, by way of approximation, $\boldsymbol{Y}_\tau$ is replaced by its actual value Y_τ, which must anyhow be known for the prediction to be made. But when the same view is taken for prediction two periods ahead the expected predictor error is

$$E(\boldsymbol{f}_{\tau+2}) = Y_\tau \operatorname{var}(\boldsymbol{b}) + \operatorname{cov}(\boldsymbol{a}, \boldsymbol{b}) \tag{9}$$

which is only fortuitously zero.

The result is that unbiased estimators of α and β do not lead to unbiased predictors for both one and two periods ahead, whatever their other properties, and conversely that unbiased predictors for both periods require biased estimators.

We conclude (i) that the requirements of estimation and prediction may differ, (ii) that the requirements of prediction may moreover vary with the predictand and (iii) that unbiasedness is an awkward condition for the predictors considered here. This leads us to compare the predictive performance of various methods rather than their merits as estimators in a simulation study where we prefer square error criteria to unbiasedness.

2 THREE METHODS OF PREDICTION

We consider three methods of predicting one and two periods onward from the end of a time-series $\boldsymbol{Y}_t$, $t = 0, 1, 2, \ldots, T$. For the simple autoregressive model

(1) the predictions are

$$\hat{Y}_{T+1} = \alpha_1 + b_1 Y_T, \tag{10}$$

$$\hat{Y}_{T+2} = a_2 + a_2 b_2 + b_2^2 Y_T, \tag{11}$$

and the prediction procedures differ only in the values assigned to (a_1, b_1) and (a_2, b_2).

I One period least squares. This is the common case where both (a_1, b_1) and (a_2, b_2) are the least squares estimates (a_0, b_0) obtained by minimizing

$$S_1(a, b) = \sum_{t=1}^{T} (Y_t - a - bY_{t-1})^2. \tag{12}$$

II Mixed prediction. In this case we retain (a_0, b_0) for (a_1, b_1); for (a_2, b_2) however, other values are used, *viz.* the two-period least squares estimates (a_*, b_*) that minimize

$$S_2(a, b) = \sum_{t=2}^{T} (Y_t - a - ab - b^2 Y_{t-2})^2. \tag{13}$$

This distinct two-period prediction method, that has been suggested by analogy to the one-period procedure, occasionally implies complex values for (a_*, b_*). The DYN method, that was recently investigated by Fair [6] on its predictive performance in a more complicated model, is equivalent to this procedure.

III Mixed estimation. We return to the use of a single set $(a_., b_.)$ for both (a_1, b_1) and (a_2, b_2); it is obtained by minimizing

$$S(a, b) = S_1(a, b) + S_2(a, b). \tag{14}$$

In the formulation above the computation of the coefficients is straightforward for methods I and II; but to minimize (14), two quadratic equations in (a, b) must be solved by numerical methods. An iterative algorithm was used, based on the gradient method, i.e. the generalised Newton-Raphson process, but also incorporating some devices from the revised scoring method (see Vandaele and Chowdhurry [14]) in order to ensure a minimum solution.

The prediction methods for equation (2), which includes an exogenous variable, are defined in the same way; there is no need to give the formulae. But here method II already requires the use of iterative methods since the minimum conditions for the analogue of (13) are nonlinear in the coefficients a_*, b_* and c_*.

3 THEORETICAL CONSIDERATIONS

Although our main concern is with prediction, the three procedures just described do yield coefficients that can be regarded as estimates of (α, β) and so

we shall present some of our results in this respect. Moreover an analysis of these estimators is at the basis of an investigation into the properties of the various predictors.

First we recall that the standard assumptions of least squares do not apply in autoregressive equations like (1) or (2); for a clear demonstration see Malinvaud [9]. As a result neither $(\boldsymbol{a}_0, \boldsymbol{b}_0)$ nor its counterpart $(\boldsymbol{a}_0, \boldsymbol{b}_0, \boldsymbol{c}_0)$ for model (2) are best unbiased, or even unbiased.

Malinvaud [9] (pp. 540–548) proves that under rather weak conditions they are asymptotically unbiased. But in practice we have only a limited number of observations at our disposal, say at most 40. Thus we are especially interested in the small sample behaviour of the estimators. Some important results have been established, mainly for the stationary AR(1) model without a constant term

$$\boldsymbol{Y}_t = \beta \boldsymbol{Y}_{t-1} + \boldsymbol{U}_t \tag{15}$$

where $\boldsymbol{U}_t$ is assumed to be normally distributed. In this model Hurwicz [7] found a downward bias for $\boldsymbol{b}_0$ as is shown by

$$E(\boldsymbol{b}_0) = \beta\left(1 - \frac{2}{T}\right) + 0(T^{-2}). \tag{16}$$

Kendall and Stuart [8] (chapter 48) present results for various estimators of the serial correlation, which, for the appropriate lags, are almost identical to $\boldsymbol{b}_0$ and $\boldsymbol{b}_*^2$. The variance of $\boldsymbol{b}_0$ and $\boldsymbol{b}_*^2$ appears to be independent of the variance of the disturbances:

$$\operatorname{var}(\boldsymbol{b}_0) \simeq \frac{1 - \beta^2}{T + 1} + 0(T^{-2}) \tag{17}$$

and

$$\operatorname{var}(\boldsymbol{b}_*^2) \simeq \frac{(1 - \beta^2)(1 + 3\beta^2)}{T + 1} + 0(T^{-2}). \tag{18}$$

Let $\mathrm{DVAR}_{\mathrm{I,II}}$ be the variance of the predictor error of method I in the two period prediction minus the corresponding variance for method II. If we assume that

$$E(\boldsymbol{b}_0) = \beta + 0(T^{-1}) \tag{19}$$

$$E(\boldsymbol{b}_*^2) = \beta^2 + 0(T^{-1}) \tag{20}$$

and again Y_T is regarded as non-random in the prediction, it can be calculated that for (15)

$$\mathrm{DVAR}_{\mathrm{I,II}} \simeq \frac{-(1 - \beta^2)^2 Y_T^2}{T + 1} + 0(T^{-2}). \tag{21}$$

$DVAR_{I,II}$ is negative for all $|\beta| < 1$ which suggests that method II will not perform very well as compared with method I, when a series is really generated by that simple AR(1) model.

There are also a great number of simulation studies on estimation and prediction with autoregressive models. We will briefly review two of them.

Malinvaud [10] examined models of the forms (1), (2) and (15); $\boldsymbol{b}_0$ was indeed downward biased and its distribution appeared to be skew with a long tail of small values.

Orcutt and Winokur [12] found that for model (1) a correction of $\boldsymbol{b}_0$ for its bias had a favourable effect on $(\boldsymbol{b}_0 - \beta)^2$ only if $|\beta| > 0.7$. Presumably this kind of correction will be of little value in improving predictions, the more so as the variance of the predictor error is independent of the bias of $\boldsymbol{b}_0$. This is illustrated by the fact that only conditions like (19) and (20) and not like (16) are necessary to calculate the variance up to $0(T^{-2})$.

All studies show a considerable bias in least squares estimators of all parameters of autoregressive models. Its size varies with the parameters of the model.

4 DESIGN OF THE DATA

The artificial time series of the present analysis have all been generated by particular specifications of the general formula

$$\boldsymbol{Y}_t = \alpha + \beta_1 \boldsymbol{Y}_{t-1} + \beta_2 \boldsymbol{Y}_{t-2} + \beta_3 \boldsymbol{Y}_{t-3} + \gamma X_t + \boldsymbol{U}_t, \tag{22}$$

where the $\boldsymbol{U}_t$ are independent $N(0, \sigma^2)$ variates throughout.

The exogenous variable X is introduced in the data for equation (2) only; two series from Malinvaud [10] are used. Series A shows a trend, and series B a cyclical movement (see appendix).

Important changes in the data are obtained by varying β_1, β_2 and β_3. The single lag equations (1) and (2) that we consider are appropriate only if $\beta_2 = \beta_3 = 0$; in all other cases they represent misspecifications. Moreover the single lag models with autocorrelated disturbances discussed in other studies like Malinvaud [10] can easily be rewritten as autoregressive models with $\beta_2 \neq 0$ and independent disturbances.[2]

Apart from β_1, β_2 and β_3 we also vary the other parameters α, γ and σ.

The starting values Y_s of Y are always equated to the solution of

$$Y_s = \alpha + \beta_1 Y_s + \beta_2 Y_s + \beta_3 Y_s + \gamma X_0 \tag{23}$$

2 Generally a number of different specifications of AR(m–n) models with AR(n) disturbances will generate the same series that can uniquely be represented by one single specification of an AR(m) model with white noise.

TABLE 1 – CHARACTERISTICS OF SIMULATED TIME SERIES

code	β_1	β_2	β_3	α	γ	type of series X	σ	T
			10 time series generated for each					
1	0.8	0	0	0	–	–	5	20
2	0.8	0	0	0	–	–	10	20
3	0.8	0	0	100	–	–	20	20
4	0.8	0	0	100	–	–	10	40
5	0.8	0	0	100	–	–	10	10
6	0.4	0	0	0	–	–	5	20
7	0.4	0	0	100	–	–	10	20
8	0.2	0	0	100	–	–	10	20
9	0.6	0.2	0	0	–	–	20	20
10	0.4	0.4	0	100	–	–	20	20
11	0.2	0.2	0	100	–	–	20	20
12	0.2	0.2	0	0	–	–	10	20
13	0.4	0.2	0.2	100	–	–	20	20
14	0.6	−0.2	0	100	–	–	20	20
15	0.8	0	0	100	0.5	A	40	20
16	0.8	0	0	100	2.5	B	40	20
17	0.4	0	0	100	0.5	A	20	20
18	−0.4	0	0	200	1	A	20	20
19	0.6	0.2	0	100	2.5	A	40	20
20	0.4	0.4	0	100	1	A	40	20
21	0.4	0.4	0	100	1	B	40	20
22	0.6	−0.2	0	100	1	B	20	20
			40 time series generated for each					
23	0.8	0	0	100	–	–	10	20
24	0.4	0	0	0	–	–	10	20
25	0.6	0.2	0	100	–	–	20	20

so that Y_s is the equilibrium value of the model that is obtained for $X_t = X_0$ for all t.[3]

Altogether these variations lead to 25 different specifications as shown in table 1. This, of course, is only a modest and rather arbitrary sample from all possible combinations. For most of these specifications 10 time series were generated, but in three cases we prepared 40 time series. Almost all series consist of 20 observations, which means that up to 25 values of Y_t must be gener-

3 Only stationary series are generated.

ated, *viz.* three (equal) starting values at most, 20 values used in estimation and fitting and two subsequent values for testing the predictions.

Unfortunately the data are not as numerous as in some other studies but there are limits to the amount of computation that is acceptable. Since each time series is subject to three alternative methods of estimation and prediction which involve at least one iterative procedure, the computations are already very extensive. As can be seen from table 1 they bear on 340 time series, distributed according to their major characteristics in table 2.

TABLE 2 – DISTRIBUTION OF ARTIFICIAL TIME SERIES ACCORDING TO THEIR MAJOR CHARACTERISTICS

	NP: $\gamma = 0$, no predetermined variable X		*PX*: predetermined variable X	Total number of series
	10 obs.	40 obs.	10 obs.	
NM: $\beta_2 = 0$ no misspecifications	80	80	40	200
M: $\beta_2 \neq 0$ misspecifications	60	40	40	140
Total number of series	140	120	80	340

5 DISCUSSION OF THE ESTIMATES

To begin with the various estimates of β_1 will be compared. The results for b_* permitted no clear conclusion, but those for b_0 and $b_.$ do, and they are shown in table 3. The estimated coefficients have been expressed as a percentage of the true β, and the average of these values over all the simulations concerned is shown in each entry of the table.

TABLE 3 – ESTIMATES AS A PERCENTAGE OF TRUE β_1

b_0 (method I)

	NP		*PX*	Total
	10	40	10	
NM	62	74	83	71
M	69	78	106	82
Total	65	75	94	75

$b_.$ (method III)

	NP		*PX*	Total
	10	40	10	
NM	60	71	80	68
M	79	77	113	88
Total	68	73	96	76

These values confirm the results of both analysis and of earlier studies to the effect that the least squares estimator b_0 underestimates β_1, excepted the case of $\gamma \neq 0$ along with misspecifications.

The corresponding average values of the estimates of α are given in table 4 for those series where $\alpha \neq 0$.

TABLE 4 – ESTIMATES AS A PERCENTAGE OF TRUE α

a_0 (method I)

	NP		PX	Total
	10	40	10	192
NM	184	220	175	192
M	227	268	268	254
Total	203	244	221	222

$a_.$ (method III)

	NP		PX	Total
	10	40	10	
NM	190	225	172	195
M	218	271	240	243
Total	203	248	206	218

Underestimation of β appears to go together with overestimation of α. This is no surprise in the case of the common least squares method:

$$\overline{Y} = a_0 + b_0 \overline{\overline{Y}} \tag{24}$$

with

$$\overline{Y} = \sum_{t=1}^{T} Y_t/T \quad \text{and} \quad \overline{\overline{Y}} = \sum_{t=1}^{T} Y_{t-1}/T$$

while for the generated series

$$\overline{Y} = \alpha + \beta \overline{\overline{Y}} + \overline{U} \tag{25}$$

$$(\overline{U} = \sum_{t=1}^{T} U_t/T)$$

so that

$$a_0 = \alpha + (\beta - b_0)\overline{\overline{Y}} + \overline{U}. \tag{26}$$

In those series where $\alpha \neq 0$, $\overline{U}$ is relatively small. As a result a_0 and b_0 tend to be inversely related. This effect is also apparent in the joint frequency distributions of $(a_0 - \alpha)$ and $(b_0 - \beta)$.

It is noticeable that overestimation of α is even greater in the case of higher order lags: positive values of β_2 (and β_3) increase $\overline{\overline{Y}}$ and this seems to be partly reflected in the value of the constant term.

Table 5 shows the estimates of γ (provided $\gamma \neq 0$). Sometimes a substantial difference from the true value can be observed but there is no systematic pattern.

As was found in other studies, the least squares method somewhat underestimates the value of the standard deviation σ in model (1), but not in model (2). The percentual standard deviation of the coefficient b_0 appeared indeed to

TABLE 5 – AVERAGE VALUES OF THE ESTIMATES OF γ (NOT PERCENTAGES!)

N° of simulation	Several lags	Type of series X	γ	c_0 method I	c_* method II	$c_.$ method III
15	no	A	0.5	0.83	0.97	0.80
16	no	B	2.5	2.40	2.39	2.39
17	no	A	0.5	0.64	0.78	0.69
18	no	A	1	0.99	0.93	0.98
19	yes	A	2.5	2.81	2.66	2.70
20	yes	A	1	1.55	1.22	1.40
21	yes	B	1	0.62	0.58	0.60
22	yes	B	1	1.31	1.39	1.35

be almost independent of α. Three times 10 series were made with $\alpha = 100$, $\beta_1 = 0.8$, $\beta_2 = \beta_3 = \gamma = 0$, and σ respectively 2, 10 and 30. The mean values of the estimates are resp. 36.6, 47.7 and 35.9% of the b_0 (and 49.9, 50.1 and 69.3% of the a_0).

Finally the Durbin-Watson statistic [5] is computed. Though this test is strictly not applicable in the case of lagged variables, in practice it is nearly always employed (see Cramer [4], page 199 and table 19), and models registering significant autocorrelation of the disturbances are rejected. The performance of the test was rather poor: altogether there were only 3 5% significant values among the 340 simulations, and even none of them occurred when there were true misspecifications.

In conclusion the differences between the alternative estimates are disappointingly small, much smaller than between the estimates and the real values.

6 PREDICTION

For any single simulated time series each of the three procedures yields predictions one and two periods ahead of the last observation included in their estimation. Since they have all been fitted by minimizing some sum of squares, it would seem proper to assess their predictive performance also by squared prediction errors, for example by

$$f_{T+1}^2 + f_{T+2}^2. \tag{27}$$

In practice, however, this criterion does not bring out the relative merits of various prediction methods, since the prediction errors largely reflect the contemporaneous disturbances of the prediction period. Take, for instance, the simple case of a time series generated by

$$\boldsymbol{Y}_t = \alpha + \beta \boldsymbol{Y}_{t-1} + \boldsymbol{U}_t \tag{28}$$

combined with a prediction scheme of the form

$$\hat{Y}_{\tau+1} = a + bY_{\tau}. \tag{29}$$

The predictor errors are

$$f_{T+1} = (a - \alpha) + (b - \beta)Y_T + U_{T+1} \tag{30}$$

$$f_{T+2} = (a - \alpha) + (ab - \alpha\beta) + (b^2 - \beta^2)Y_T + \beta U_{T+1} + U_{T+2}. \tag{31}$$

In both cases, the terms U_{T+1}, and U_{T+1} and U_{T+2} respectively tend to predominate quite heavily in the square of the predictor errors, as can be verified from their expected values

$$E(f_{T+1}^2) = E\{(a - \alpha) + (b - \beta)Y_T\}^2 + \operatorname{var}(U) \tag{32}$$

$$E(f_{T+2}^2) = E\{(a - \alpha) + (ab - \alpha\beta) + (b^2 - \beta^2)Y_T\}^2 + (1 + \beta^2)\operatorname{var}(U). \tag{33}$$

Clearly, the systematic part of these expressions is as a rule much smaller than the terms representing residual variance.

By itself this would not affect the comparison between predictions for a given time series since the same residual component is present with all predictions. However, if the comparison bears on a number of simulations, as is bound to be the case, the residual component adds greatly to the variation between series and thus seriously reduces the possibility of detecting significant differences. This fact has been noticed by Malinvaud [10].

While this difficulty could be overcome by increasing the number of simulations we have preferred to avoid it altogether by omitting the contemporaneous disturbances from the simulated values for periods $T + 1$ and $T + 2$, and thus from the prediction errors. As all the further tests are based on pairwise comparisons of different predictions for the same series this does not affect the relative position of each method while it greatly improves the chances of detecting significant differences from a small number of trials. More precisely, the predictive performance of any method on each trial is gauged by

$$EPS = f_{T+1}^{*2} + f_{T+2}^{*2} \tag{34}$$

with

$$f_{T+1}^{*} = f_{T+1} - U_{T+1} \tag{35}$$

$$f_{T+2}^{*} = f_{T+2} - \beta U_{T+1} - U_{T+2}. \tag{36}$$

7 DISCUSSION ON THE PREDICTION RESULTS

Since the distribution of *EPS* is unknown, the predictive performance of the three methods will be compared by distribution-free tests of the *EPS* (one for each method), obtained for each simulation. It should be noted that we do not compare the squared prediction errors separately but only their sum because method I and method II merely differ in the two period predictions and method III has especially been designed for predicting one and two periods together.

First the null hypothesis of equal predictive performance of the three methods is simultaneously tested by Friedman's test (see Conover [3], pp. 265–270). Table 6 shows the results.

TABLE 6 – RESULTS OF FRIEDMAN'S TEST

	NP		*PX*	Total
	10	40	10	
NM	686 (480)	1466 (480)	38 (240)	3426 (1200)
M	134 (360)	518 (240)	18 (240)	806 (840)
Total	529 (840)	3512 (720)	98 (480)	6686 (2040)

critical value ($\alpha = 0.05$) in parentheses.

The null hypothesis is rejected if the test statistic is greater than the critical value.

It is also checked how often each of the methods yielded the lowest *EPS*. The score is for method I: 171, for method II: 100 and for method III: 69. The poor performance of the latter is not surprising: method III lies in between the two others and it often does so in performance as well.

Since method II and III are put foreward as alternatives to the standard least squares method I, the null hypothesis of equal predictive ability is tested for the two pairs (method II, method I) and (method III, method I). The results of the sign-test (see Conover [3], pp. 121–126), a qualitative test, are given in table 7.

The *EPS* vary substantially. Yet a quantitative pairwise comparison between the prediction methods can be made by the following indices

$$I \text{ (method II, method I)} = \frac{EPS \text{ II}}{0.5 \cdot (EPS \text{ I} + EPS \text{ II})} \cdot 100 \tag{37}$$

This index is symmetric to the effect that

$$\frac{I \text{ (method II, method I)} + I \text{ (method I, method II)}}{2} = 100. \tag{38}$$

TABLE 7 – SIGN-TEST ON (*EPS* II – *EPS* I)[4]

	NP			*PX*	Total
	10		40	10	
NM	(30) 31 (50)		(30) 29 (50)	(13) 23 (27)	(85) 83 (115)
M	(21) 26 (39)		(13) 10 (27)	(13) 22 (27)	(57) 58 (83)
Total	(57) 57 (83)		(48) 39 (72)	(30) 45 (50)	(161) 141 (189)

SIGN-TEST ON (*EPS* III – *EPS* I)[4]

	NP			*PX*	Total
	10		40	10	
NM	(30) 35 (50)		(30) 31 (50)	(13) 19 (27)	(85) 85 (115)
M	(21) 25 (39)		(13) 13 (27)	(13) 21 (27)	(57) 59 (83)
Total	(57) 60 (83)		(48) 44 (72)	(30) 40 (50)	(161) 144 (189)

two sided test: critical values ($\alpha = 0.05$) in parentheses.

I (method III, method I) is calculated in the same way as (37).
The mean values of these indices are given in table 8.

TABLE 8 – *I* (METHOD II, METHOD I): MEAN VALUES

	NP		*PX*	Total
	10	40	10	
NM	108.7	114.2	102.8	109.7
M	103.8	108.7	102.5	104.8
Total	106.6	112.4	102.7	107.7

I (METHOD III, METHOD I): MEAN VALUES

	NP		*PX*	Total
	10	40	10	
NM	100.6	107.1	99.3	102.9
M	107.2	101.5	102.3	104.2
Total	103.4	105.2	100.8	103.4

4 The test statistic indicates how often method II (resp. method III) is better than method I. If it lies in between the critical values the null hypothesis is not rejected.

All values but one are greater than 100. This indicates a good performance of the standard least squares method. These mean values do not differ very much from 100, the alternative predictions for the individual simulations, however, often show considerable differences.

Now the null hypothesis of equal predictive ability can be tested for the two pairs by the signed-rank-test (see Conover [3], pp. 206–215), which is more powerful than the sign-test. The index

$$I_r \text{ (method } i\text{, method I)} = I \text{ (method } i\text{, method I)} - 100 \tag{39}$$

is used. These I_r's are ordered in absolute value and the rank numbers are added with the appropriate sign. In large samples the test statistic can be approximated by a standard normal distribution and these values are given in table 9.

TABLE 9 – SIGNED-RANK-TEST STATISTICS, NORMAL APPROXIMATION, TWO SIDED TEST
(CRITICAL VALUES FOR $\alpha = 0.05$: -1.96 AND $+1.96$)

	(method II, method I)					(method III, method I)			
	NP		*PX*	Total		*NP*		*PX*	Total
	10	40	10	3.31		10	40	10	
NM	2.14	3.10	−0.30	3.31	*NM*	0.75	2.44	−0.01	2.03
M	0.83	2.03	−0.32	1.46	*M*	1.86	1.08	−0.35	1.57
Total	2.22	3.71	−0.44	3.50	Total	1.76	2.63	−0.21	2.62

From these values we can conclude that method I is significantly better than method II if no exogenous variable is introduced in the model: so that the theoretical result (21) is confirmed. Method III is also worse, but to a smaller extent.

On the other hand no large differences were found in case of model (2) and here the three alternative methods appear to perform equally well.

8 HISTORICAL TIME SERIES

The predictive performance of the three methods for model (1) is also assessed on the basis of a number of historical (as opposed to artificial) time series. These are selected from larger collections of U.S. and Dutch time series so as to discard all items for which model (1) would be normally rejected, or, more precisely, by retaining only those series that upon the application of least squares

- show no 'significant' autocorrelation of the disturbances by the Durbin-Watson statistic
- yield estimates (a_0, b_0) larger in absolute value than their standard error.

No attention is paid to the question whether a plausible (economic) motivation could be found for using model (1).

The above criteria are applied to 59 series of the United States from Tinbergen [13] and to 73 postwar series for the Netherlands [2]. The first are chosen because Orcutt [11] already investigated their autoregressive nature. The latter are selected on the requirements that they show some cyclical movement and not too much trend. From the American series 30 meet the two conditions given above, and from the Dutch 49. The Durbin-Watson statistic remarkably often appeared to be (too) low; much more frequently than in the simulation study it suggested the presence of misspecifications. The reference period of the U.S. series is 1919–1930 ($T = 11$) and predictions are made for 1931 and 1932. In the case of the Dutch series: reference period 1946–1965 or 1947–1966 ($T = 19$); predictions 1966, 1967 or 1967, 1968.

Table 10 shows the distribution of the values of b_0.

TABLE 10 – ESTIMATES OF b_0 IN HISTORICAL TIME SERIES

b_0	<0	0–0.33	0.33–0.66	0.66–1	>1	Total
U.S.	1	1	12	14	2	30
Netherlands	2	5	28	14	0	49

The prediction errors are submitted to the same statistical tests as in the simulation study. The results are shown in table 11.

TABLE 11 – CHARACTERISTICS OF PREDICTION ERRORS OF HISTORICAL SERIES

Friedman's test ($\alpha = 0.05$)			
U.S.	628 (180)		
Neth.	86 (252)		
	method I best	method II best	method III best
U.S.	24	4	2
Neth.	16	16	17 (!)
Sign-test ($\alpha = 0.05$)			
	II better than I	III better than I	
U.S.	(12) 6 (18)	(12) 6 (18)	
Neth.	(17) 27 (32)	(17) 28 (32)	

	I (method II, method I)	I (method III, method I)
U.S.	102.6	104.1
Neth.	99.7	101.6

Signed-rank-test, standard normal approximation (critical values $\alpha = 0.05$: -1.96 and $+1.96$)

	(method II, method I)	(method III, method I)
U.S.	2.99	3.73
Neth.	-1.26	-0.52

For the American series the hypothesis of the three methods being equal is rejected by all tests. The standard least squares method is significantly the best. However the choice of the years to predict is very unfortunate: 1931 and 1932, just after the great crisis. The prediction errors are so enormous that we should be wary of drawing conclusions from these series.

In the Dutch series the performance of method II and III is certainly not inferior to that of method I. A different conclusion than from the simulation study!

9 FINAL REMARKS

One of the Dutch series, the number of divorces yearly per 10000 married couples, shows such a magnificent least squares fit that it would mean a cheerful day to anyone engaged in econometric research:

$$\underset{}{Y_t} = \underset{(11.2)}{6.03} + \underset{(3.5)}{0.726}Y_{t-1} \qquad \underset{2.06}{D.W.S.} \tag{40}$$

(percentage standard errors in parentheses)

Yet we have to draw the paradoxical conclusion that this model strictly cannot be right, because the standard errors are too low! They do not agree with (17) or with corresponding values found in the simulation study. Hence we must either conclude that the model must be rejected, since the results are significantly at variance with it, or that we have met a rare chance event. The latter interpretation is in order in so far as the example given is itself selected as the 'best' least squares fit of some 100 historical series.

Yet this case illustrates a serious difficulty of empirical econometrics. After much trial and error nearly always a model can be found that fits the observations well, or (as in the present case) 'suitable' observations can be found for a certain model. This leaves the question unanswered whether such a model really describes well the relation and yields good predictions.

The question, what method to use if two periods ahead must be predicted by an autoregressive model with one lag is solved to some degree. The usual least squares method performs relatively well, and is definitely superior to method II (identical to DYN) in the simple AR(1) model. For the model with a predetermined variable in the simulation study and for the historical Dutch time series no significant differences between the three methods could be found.

In contrast with these remarks Fair [6] concluded the DYN procedure to be rather successful. Though he looked at within sample-, and not at ex ante predictions, it may be true that in more complicated autoregressive models, with a number of predetermined variables, gain can be achieved by this prediction procedure. But our results are reassuring in that there is no immediate need for reestimating all existing autoregressive models to get better predictions.

APPENDIX

The value of the exogenous variable X, copied from Malinvaud [10].

t	A	B	t	A	B
0	80	110	12	100	88
1	86	105	13	116	92
2	78	82	14	125	115
3	72	80	15	130	139
4	62	78	16	145	90
5	62	82	17	153	80
6	68	88	18	138	74
7	74	102	19	139	78
8	84	107	20	144	86
9	88	95	21	156	105
10	84	101	22	167	138
11	91	92			

REFERENCES

[1] BOX, G. E. P. and G. M. JENKINS, *Time Series Analysis*, San Francisco, 1970.

[2] C.B.S., 1899–1969, *Zeventig jaren statistiek in tijdreeksen*, 's-Gravenhage, 1970.

[3] CONOVER, W. J., *Practical Nonparametric Statistics*, New York, 1971.

[4] CRAMER, J. S., *Empirical Econometrics*, Amsterdam, 1969.

[5] DURBIN, J. and G. S. WATSON, 'Testing for Serial Correlation in Least Squares Regression,' *I Biometrika*, XXXVII (1950), pp. 409–428, *II Biometrika*, XXXVIII (1951), pp. 159–178.

[6] FAIR, R. C., 'A Comparison of Alternative Estimators of Macro-Economic Models,' *International Economic Review*, XIV (1973), pp. 261–277.

[7] HURWICZ, L., 'Prediction and Least Squares,' Chap. XV in T. C. Koopmans (ed.), *Statistical Inference in Dynamic Economic Models*, New York, 1950, pp. 365–383.
[8] KENDALL, M. G. and A. STUART, *The Advanced Theory of Statistics*, Vol. III, 2nd. edn., London, 1968.
[9] MALINVAUD, E., *Statistical Methods of Econometrics*, Second Revised Edition, Amsterdam, 1970.
[10] MALINVAUD, E., 'Estimation et prevision dans les modeles economiques autoregressifs,' *Revue Inst. Int. de Stat.* 29: 2, (1961), pp. 1–32.
[11] ORCUTT, G. H., 'A Study of the Autoregressive Nature of the Time Series used for Tinbergen's Model of the Economic System of the United States, 1919–1932,' *Journal of the Royal Statistical Society*, Series B, (1948), pp. 1–45.
[12] ORCUTT, G. H. and H. S. WINOKUR, 'First Order Autoregression: Inference, Estimation, and Prediction,' *Econometrica*, XXXVII (1969), pp. 1–14.
[13] TINBERGEN, J., *Statistical Testing of Business Cycles Theories, Vol. II: Business Cycles in the U.S.A.*, 1919–1932, League of Nations, Geneva, 1939.
[14] VANDAELE, W. H. and S. R. CHOWDHURRY, *A Revised Method of Scoring*, Institute of Economics, Tilburg, 1970.

Summary

PREDICTION OF TWO PERIODS BY SIMPLE AUTOREGRESSIVE MODELS WITH ONE LAG

Three different methods are compared by their ability to predict two periods ahead in simple autoregressive models with one lag. In this study both artificial and historical time series are used. In spite of intuitive objections the usual least squares method performs relatively well. Moreover attention is paid to the estimation results, as they provide some links with other studies of the autoregressive model.